



EST:

gi 10725490	/dataset=dbest /taxon=96...	858	0.0
gi 10947399	/dataset=dbest /taxon=96...	846	0.0
gi 9121897	/dataset=dbest /taxon=9606...	846	0.0
gi 13280819	/dataset=dbest /taxon=96...	846	0.0
gi 13287907	/dataset=dbest /taxon=96...	833	0.0
gi 13286505	/dataset=dbest /taxon=96...	831	0.0
gi 8150776	/dataset=dbest /taxon=960...	815	0.0
gi 5936410	/dataset=dbest /taxon=9606 ...	726	0.0
gi 6868875	/dataset=dbest /taxon=9606...	726	0.0
gi 6868872	/dataset=dbest /taxon=9606...	726	0.0

**EXPRESSION INFORMATION FOR MODULATORY USE:**

library source:

gi|10725490|adult adrenal gland  
gi|10947399|mammary gland  
gi|9121897|retinoblastoma  
gi|13280819|adenocarcinoma cell line  
gi|13287907|retinoblastoma  
gi|13286505|embryonal carcinoma, cell line  
gi|8150776|adult uterus  
gi|5936410|adult uterus  
gi|6868875|adult head\_neck  
gi|6868872|adult head\_neck

Tissue Expression:

Human leukocyte

# Isogen 2:

1 GGAGGATTAA TCAGTGACAG GAAGCTTCTT CTTCGGAGC GGTGACCAGC  
51 TGTGGTTAGT AGAGGCTCAG CACGGGTAGT CCCAGGAGTC TTTCGGATT  
101 CTTGCTTACT GGTACCCAC CTGCTCTTGC CATGAGGCAC CTTGGGGGCT  
151 TCCTCTTCTT TCTGGGGGTC CTGGGGGCTT TCAGTGAGAT GTGTGAATA  
201 CCAGAGATGG ACAGCCATCT GGTACAGAAG TTGGGGCAGC ACCTCTTACC  
251 TTCGATGAC CGGCTTTCCC TGCACCACTT GAACCCAGC ATCTATGTGG  
301 GCCTACCTCT CTCCAGTCTG CACCTTGGGA CCAAGGAAGA CCTCTACCTG  
351 CACAGCTTCA TGCTTGGTTA CCACCACTGC CTCCACGGT CTCTCTCAG  
401 CGAGGATGAC GGTGACTGCC AGGCCAAGCC TTCCATCGGC TAAGTGGCCC  
451 TCTACCTCT CTGCTCTCAGA GCCAATCGGC ATGATCACA GGGCCACCCC  
501 CACACTAGT ACTACCACTA TGGCTCTGGC ATTCTGCCCC TGCTCTCCA  
551 CCAAGAGCAG GTCCATGACA CCGTGGTGA CAACTTCTG TAAGTGTGG  
601 AAGCTTTCTA CCAGGGCCAC CATCTCTGTG ACAGAGTACG CATGGCAGGC  
651 TTGCTATT TA CTCTCTGAA CCGCTCAAAC TTCAACTCTG CTCTAGACA  
701 AAGGATCA TC ATGCCATCA GAAGCTGCG AGAGGATAC TTGAAGGCC  
751 AAGCTCCC TA GGGCCACTTT GGAATGTCT ACAACACCC ATGCTATTA  
801 CAGTCTCT TA GACCTCCCC CATCTCTGG GCAAAATGG GAAGGCTATG  
851 TCTAGAGG AG AGGCTGCTT TGTCTCTAG TCTCTAGAT GAGCTTTCT  
901 AAGAGCTT TA ATGATTTCT CATCTCTGC CCGTCTGAA CCAAGAGCC  
951 TAAATGAGC TGATCTTCCC AAGCTCTCTG GCAACAGAG TATGTTTGA  
1001 AAGCTCTT TA AGACCTTCT CTAGACCCA AAGATCATC AGTGTACGC  
1051 TCAAGTGT TA TATCTCTTG GGGCTGACA GACAGTCTAT CAGTCTCTG  
1101 GCTGCTCT TA GGTGGAAGA TGTCTGAAAG AAGCTCTAT GAGTGAAGG  
1151 AATTAATAT TAAACACAGG CTTCTCTTCT AGGCTCTAT TAACTCTG  
1201 TATGAGGA TA AGCGGCCGGA AAAAGGGAGT TCTCTCACT TCTCTAGAC  
1251 TCAAGATC TA TACTGTTGCA AGGTATTGCT GACTACCAAC CCAAGATGG  
1301 AAGAGCTT TA GAGCTGAGGC GGTATTCTG GTAACCTCT GAGTCTCTCA  
1351 TCTAGACA TA TCGCACACT GGTAGCTCT CTATCTCTT TCTCTATCTC  
1401 CTAGAGAG TA GAAGCTGCT GAGCTCTCTG CCACTCTCT TCTCTATCTC  
1451 GAGAGCTC TA TGGGATCAC TCGAGCCACA AGCTCTCTG GAGCTCTATA  
1501 CAGAGCTC TA TCTGAGACA AGAGCCAAAG CATCTCTCT GAGAGCTCT  
1551 TCTGAGAA TA TGTGGCCAGC TGTCTCTCT AGGTCTCTCA TGAAGGCCAC  
1601 CAGATCTCT TA TATGGGCATG AAGATCTCT GACTCTCTG CAAAGAACGG  
1651 AAGCTCTA TA TCGAGGTGT TGTCAAGACC ACTCTCTCT TGTCTGGGT  
1701 CTCTAGAG TA GCTCTCTCA GCGGGGGGG TATGCTCTG AGCTCTCTC  
1751 TCACTCTCT TA GTTAGAGTG GCAGCTCTCA GCTCTCTCT GCACTCTAGC  
1801 TGTGAGAG TA TAGCAGGGC TGTCTCTCTG CTGCTCTCT CAAATATTA  
1851 GATCTCTCT TA TGTGAAAA TAAAAAAA AAAAAAAA AAAAAA  
(SEQ ID NO:1)

## FEATURES:

5'UTR: 1 - 111  
Start Codon: 11  
Stop Codon: 152  
3'UTR: 155

## Homologous proteins:

Top 10 BLAST Hits

	Score	E
CRA110800014556136 /altid=gi129316 /def=gb AAE15516.1  trans...	793	0.0
CRA110800014553190 /altid=gi1274175 /def=ref NP_009122.2  tr...	793	0.0
CRA11080001455111 /altid=gi1339201 /def=gb AAE1051.1  (L01643...	792	0.0
CRA110800014542036 /altid=gi12654675 /def=gb AAH0117.1  AAH01...	792	0.0
CRA1100001024610 /altid=gi14507409 /def=ref NP_00034.1  tran...	788	0.0
CRA1100000193132 /altid=gi1339201 /def=gb AAE1056.1  (L02647...	786	0.0
CRA1100000193132 /altid=gi17657639 /def=ref NP_056564.1  tran...	561	e-159
CRA1164000156745249 /altid=gi11968124 /def=ref NP_071979.1  tr...	554	e-156

CRA1180 0000018941 /alt-id=gi114507407 /def-ref:13-1-53.11 tran... 545 e-174  
 CRA118000004928134 /alt-id=gi114507407 /def-ref:13-1-53.11 tran... 128 1e-24

EST:

gi110125490 /dataset=dbest /taxon=96...	858	0.0
gi15936410 /dataset=dbest /taxon=9606...	835	0.0
gi16888975 /dataset=dbest /taxon=9606...	726	0.0
gi16888872 /dataset=dbest /taxon=9606...	726	0.0
gi112156937 /dataset=dbest /taxon=960...	686	0.0
gi110947399 /dataset=dbest /taxon=96...	680	0.0
gi113187907 /dataset=dbest /taxon=96...	680	0.0
gi19111897 /dataset=dbest /taxon=9606...	680	0.0
gi11210319 /dataset=dbest /taxon=96...	680	0.0
gi18110776 /dataset=dbest /taxon=960...	656	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

gi110125490| adult adrenal gland  
 gi15936410| adult uterus  
 gi16888975| adult head\_neck  
 gi16888872| adult head\_neck  
 gi112156937| adult lung\_tumor  
 gi110947399| mammary gland  
 gi113187907| retinoblastoma  
 gi19111897| retinoblastoma  
 gi113180819| adenocarcinoma cell line  
 gi18110776|

Tissue Expression:

Human hippocampus

1 MHEHVAPELE LQVLGALTEM CEIPEMDSHL VEPVQHLHP WMDRLGLEHL  
 51 NPLTYVGLRL SFLQACTKEE LYLHSLKLG YQVLISSAFS EDDGDAQKEE  
 101 SPPVIALYLL ALRANCEPVR GHFGDRLVSQ LZWFEDEKKE AIDTAAMAGL  
 151 APTCHERSNF MFCRPQRITM AIRTVREEIL KACTPEGHEG NVYJTFLAIQ  
 201 ELMTAFMRGA ELGTACLKAR VALLASLDQG AFQNALMISQ LLFVINHKTY  
 251 ITHIFFDCLA PRVMLEPAAE TIPQTQEIIS VTIQVLSLLP FYRQSISVLA  
 301 GNTVENVIKK AHELGCETYE TQASLSGPYL TSVMGKAAGE REFWQLIPDP  
 351 NTPLLQSIAD YRPKDGSETIE LELVSW  
 (SEQ ID NO:3)

#### FEATURES:

##### Functional domains and key regions:

PDOC00005 PS00005 PKC\_PHOSPH\_SITE

Protein kinase C phosphorylation site

Number of matches: 2

1	75-77	SLK
2	174-176	TVE

PDOC00006 PS00006 CK2\_PHOSPH\_SITE

Casein kinase II phosphorylation site

Number of matches: 6

1	57-59	TKED
2	90-98	SELD
3	174-177	TVEE
4	226-228	SLKD
5	249-252	TYID
6	302-305	STVE

PDOC00008 PS00008 MYFISTYL

N-myristoylation site

Number of matches: 7

1	12-14	SYLHGL
2	57-60	GLRLSS
3	86-91	GLAFSE
4	149-154	GLAFTC
5	191-194	SHVYST
6	209-214	GAELGT
7	230-235	GAFLNA

PDOC00009 PS00009 AMIDATION

Amidation site

162-165	PGRR
---------	------

#### SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.38	YES
mean S	1-18	0.949	0.46	YES

Most likely cleavage site between pos. 18 and 19: ALT-EM

#### BLAST Alignment to Top Hit:

>CRA|108000024636236 /altid=gi|298316 /def=gb|AAE25526.1|  
 transcobalamin II, PC II [human, endothelial cells,  
 Peptide, 427 aa] .org=human /taxon=9606 /dataset=nraa  
 /length=427  
 Length = 427

Score = 732 bits (1870), Expect = 0.0

Identities = 376/427 (88%), Positives = 376/427 (88%), Gaps = 51/427 (11%)

Frame = +1

```

Query: 41  NRHLNAPFLFLGVLGAALTEMCEIPEMDSHLVEKLGQHLLPWNFRLSI EHLNPSIYYVGLRL 210
Sbjct: 1  NRHLNAPFLFLGVLGAALTEMCEIPEMDSHLVEKLGQHLLPWNFRLSI EHLNPSIYYVGLRL 40

Query: 411  SLLQACTKEDLYLHSLKLGYYQQCLLGSAPFSEDDGDCQGGKPSMGQLALYLLALRANCEFVR 390
Sbjct: 61  SLLQACTKEDLYLHSLKLGYYQQCLLGSAPFSEDDGDCQGGKPSMGQLALYLLALRANCEFVR 120

Query: 341  GHFGDRLVSQLKWFLEDEKRAI----- 456
Sbjct: 121  GHFGDRLVSQLKWFLEDEKRAIGHDHRGHPHTSYYYQYGLGILALCLHQKRVHDSVVDKLL 180

Query: 457  -----DTAAMAGLAFTCLKFSNENPGRRQRITMAIRTVREEILKAQTPEGHF 597
Sbjct: 181  YAVEPFHQGHHSVDTAAMAGLAFTCLKFSNENPGRRQRITMAIRTVREEILKAQTPEGHF 240

Query: 598  GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 777
Sbjct: 241  GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 300

Query: 778  YIDLIFPDCLAPFVMLEPAAETIPQTQEIISVTLQVLSLLPPYFQSIISVLGASTVEDVLK 957
Sbjct: 301  YIDLIFPDCLAPFVMLEPAAETIPQTQEIISVTLQVLSLLPPYFQSIISVLGASTVEDVLK 360

Query: 958  KAHELGGFTYETQASLSGFPYLTSVMGKAAGEREFWQLLEDPNTELLQGIADYRPKDGETI 1137
Sbjct: 361  KAHELGGFTYETQASLSGFPYLTSVMGKAAGEREFWQLLEDPNTELLQGIADYRPKDGETI 420

Query: 1138  ELRLVSW 1158
Sbjct: 421  ELRLVSW 427
(SEQ ID NO:6)

```

# HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	829.9	8.6e-246	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	3.2	2.9	1

# Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11 [.]	1	11 [.]	3.2	2.9
PF01122	1/2	1	142 [.]	1	143 [.]	296.0	4.6e-85
PF01122	2/2	143	376 .]	197	450 .]	531.8	4.8e-156

isoform\_1:  
 1 MRHLCAEFL LQVIAALTEM CEIPENDSHL VEKLGQHLLP WNPVLSLEHL  
 51 NPGNYAURE SSLQAGTFED IYLSLMLGY QQCLLGSAPS EEDDQCQHP  
 101 SMGQLAAYLL ALRANWHDF GHPTSYYYQY GLGILADCLH QHPVHPSVVD  
 151 KLLYAVEEPH QGHHSVDTA MAGLAFTCLK RSNENPGRRQ RITKAIRTVR  
 201 EEIHKAGTPE GHFGNVYSTP LALQFLMTSP MRGAELGTAC LKARVALLAS  
 251 LQDGAQNAL MISQLLEVLN HKTYIDLIFP DCLAPRVMLE PAAETIPQTQ  
 301 EILSVTHQVL SLPPYRQSI SVLAGSTVED VLKKAHELGG FTYETQASLS  
 351 GPVITSVMGK AAGEREFWQL LRDPNTPLLQ GIADYRPKDG ETIELRLVSW  
 (SEQ ID NO:4)

#### FEATURES:

##### Functional domains and key regions:

PD000005 PS00005 PKC\_PHOSPHO\_SITE  
 Protein kinase C phosphorylation site  
 198-200 TYE

PD000006 PS00006 CK2\_PHOSPHO\_SITE  
 Casein kinase II phosphorylation site  
 Number of matches: 7

1	67-73	TKFD
2	90-93	SEID
3	147-153	SVND
4	198-201	TYE
5	250-253	SLQD
6	273-276	TYID
7	326-329	STVE

PD000008 PS00008 MYRISTYL  
 N-myristoylation site  
 Number of matches: 7

1	12-17	GVLCAL
2	57-62	GLFLSS
3	86-91	GSAFSE
4	173-178	GLAFTC
5	214-219	GNVYST
6	233-238	GAELGT
7	254-259	GAQNA

PD000009 PS00009 AMIDATION  
 Amidation site  
 186-189 PGER

PD0000426 PS000466 COBALAMIN\_BINDING  
 Eukaryotic cobalamin-binding proteins signature  
 165-178 SVITAAMAGLAFTC

#### SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.88	YES
mean S	1-18	0.949	0.48	YES

Most likely cleavage site between pos. 13 and 19: ALT-EM

#### BLAST Alignment to Top Hit:

>CFA|108000024636236 /altid=gi|298316 /def=gb|AAB25526.1|  
 transcobalamin II, TC II [human, endothelial cells,  
 Peptide, 427 aa] /org=human /taxon=9606 /dataset=nraa  
 /length=427  
 Length = 427

Score = 399 bits (2026), Expect = 0.1  
 Identities = 399/427 (93%), Positives = 399/427 (93%), Gaps = 27/427 (6%)

Query: 1 MRHLGAFLELLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL 60  
 Subject: 1 MRHLGAFLELLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL 60

Query: 61 SSLQAGTKEDLYLHSLMDPTQQCLLGSAFSEDDGDCQGGKPMGQLALYLLALRAN----- 115  
 Subject: 61 SSLQAGTKEDLYLHSLMDPTQQCLLGSAFSEDDGDCQGGKPMGQLALYLLALRAN 115

Query: 116 -----W-----HDHKGHPHTSYQYGLGILALCLHJKRVHDSVVDKLL 153  
 Subject: 121 SHKGDRLVSQLKWFLEDEKRAIGHDHKHPHTSYQYGLGILALCLHJKRVHDSVVDKLL 180

Query: 154 (AVEPFHQGHHSVDTAAMAGLAFTCLKPSNFNPGRRQRITHAIRTVEEILKAQTPEGHF 213  
 Subject: 181 (AVEPFHQGHHSVDTAAMAGLAFTCLKPSNFNPGRRQRITHAIRTVEEILKAQTPEGHF 240

Query: 214 GNVYSTPLALQFILMTSPMGAELGTACLKARVALLASLQDGAFCNALMISQLLEVLNHKT 273  
 Subject: 241 GNVYSTPLALQFILMTSPMGAELGTACLKARVALLASLQDGAFCNALMISQLLEVLNHKT 300

Query: 274 YIDLIFPICLAFFVMLEPAAETIPQTQEIISVTLQQLSLLIFYFQSIYVLAGSTVEDVLK 333  
 Subject: 301 YIDLIFPICLAFFVMLEPAAETIPQTQEIISVTLQQLSLLIFYFQSIYVLAGSTVEDVLK 360

Query: 334 KAEELGGFTYETQASLSGPYLTSMVGKAAGEREFFWQLLEDPNTELLQGIADYRPKDGETI 393  
 Subject: 361 KAEELGGFTYETQASLSGPYLTSMVGKAAGEREFFWQLLEDPNTELLQGIADYRPKDGETI 420

Query: 394 ELRLVSW 400  
 Subject: 421 ELRLVSW 427  
 SEQ ID NO: 7)

#### HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	904.3	8.6e-269	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	3.2	2.9	1

#### Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11 [.	1	11 [.	3.2	2.9
PF01122	1/2	1	115 [.	1	115 [.	241.3	1.4e-68
PF01122	2/2	117	400 .]	145	450 .]	660.5	8.7e-195







7181 TATTTAAAGG TGCTGTGTGG CAA TGTAAAT CTATTTAACT TGGGAGGGGG  
 7191 AATTAAGAAGA ATGGGTTGAA CTGGGGAGGG GAAATTTGGA GTGAGGTGAG  
 7201 AATATGCTAC TGGACTGAG CTG GGG AAC AAGAGGAAA CTCTGTCTGA  
 7301 ATGAAATTA TAAATAAATA TAAATA ATA AAAAGGAGG GGGCATATGG  
 7351 TTGAATATG CTATATATG TGGGCTGGG AATGATGAT TGGTATAGG  
 7361 AGCTTTGGA GTTTATCTT GATCTTACT CTATATCTTT ATTAATATTT  
 7401 TTGGGATAC ATCTATAC ATCTATAC AATAGGCTC TTTTGGAGCT  
 7451 TTATATGAG TATCTATG TTAAGATTT TTGGGCTTT TCAATACGG  
 7501 TGGGTCGCT CTATATATG GCTTTATAGG ATGAGTTT TCAATACGG  
 7551 TATTAATAG CTATATATG TATATCTT CAGTATAGG TCAATACGG  
 7601 TGGGAGGCTA CTATATATG TATATCTT TATATCTT TCAATACGG  
 7651 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 7701 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 7751 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 7801 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 7851 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 7901 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 7951 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8001 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8051 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8101 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8151 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8201 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8251 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8301 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8351 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8401 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8451 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8501 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8551 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8601 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8651 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8701 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8751 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8801 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8851 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8901 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 8951 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9001 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9051 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9101 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9151 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9201 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9251 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9301 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9351 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9401 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9451 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9501 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9551 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9601 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9651 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9701 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9751 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9801 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9851 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9901 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 9951 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10001 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10051 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10101 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10151 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10201 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10251 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10301 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10351 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10401 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10451 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10501 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10551 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG  
 10601 TCACTTAAA TATATCTT TATATCTT TATATCTT TCAATACGG

FIGURE 3, page 3 of 22

10751 GAGTGTGGGG AAGAAAGAG TAAATGGAGA TCGCTGGGGA TATGGCTTGT  
10752 GTGATAGGAA GTAGAGAAAT GTTGGGATGG ATGGATGGAT GGATGGATGG  
10753 ATGGATGGAT GGGGCGATGG ATGAGGAGAG ACACATTTTG GTTAAATGTA  
10801 ATACAGCATG ATAAAGGCTA GTAGAGAGAT GATCGAGGCT TTCTCTGAGA  
10851 GAGGGCTGTG AGGAGTGAAT GGGATTGTTT AATTAAAGAA GAGGAGAGAG  
10901 GCGAGGTGGA GTGATATATG GCTGTAATCG GAACACTTTC GGAGCTGGAG  
10951 GCGGGTGGCT GAGGTGAGGT GAGGAGTTTG AGACGAGGCT GGTAAATATG  
11001 GCGAAACTCG ATCTATTAAG AATACAAAAA AGTACCTGGG TGTGCTGGCG  
11051 AGTGGCTGTA AGGCGAGGTA AGCTATATAG GAGGCTGAGG CAAGAGAAATC  
11101 ACTTGAACCT GAGAGCTGGA GGTTCGATG AGCGAAGATC ATCGCAATGC  
11151 ACTCGACTCT GGGTGAAGGA GTAGAGATAT GTCTAAAAA AAAAAAATAA  
11201 AAATGAGAGG AAGGGAAGCT GAGATGCTTG GCTCGCTGCT AATATCTAG  
11251 GAGCTGCTGA AGGTGAGGTA GATGGAATGG CTGAGCTGAG GAGTCTGAGA  
11301 GAGGCTGCTG CAACATCTTG AAACTCTGCT TTTACTAAAA CAGGAAAGAT  
11351 GAGGCTGCTG TGGTCTGAGA GAGCTGTAAT CAGAGCTAAT AGGAGGCTTG  
11401 AGCGACAGAA ATCAGCTTGA GCTGAGAGAG AGAGCTGCTA GTGAGCTGAG  
11451 AGCTGCTGTA AGCTGCTGAG GCTGCTGAG AGCTGCTGAG GCTGCTGAG  
11501 GAAAAAATAA GAATCTGAG AGCTGCTGAG AGAGAGAGAG GAGGAAAGAA  
11551 AGAGAGAGAG GTGAGAGAG TTTGCTGAG TTTGCTGAG TTTGCTGAG  
11601 AAGAGCTGAG ATCTGAGAG ATGAGCTGAG GTGAGAGAG GAGGAAAGAA  
11651 GAGGAGCTGA AGGCTGAGAG GTAGAGAGAG TGGCTGCTGA AAGAACTGAT  
11701 GTTCTGAGAG GAGCTGCTGA GAGCTGCTGA TGGCTGCTGA GTGAGAGAG  
11751 GAGAGCTGAG AGAGAGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
11801 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
11851 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
11901 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
11951 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12001 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12051 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12101 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12151 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12201 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12251 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12301 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12351 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12401 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12451 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12501 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12551 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12601 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12651 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12701 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12751 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12801 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12851 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12901 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
12951 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13001 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13051 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13101 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13151 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13201 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13251 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13301 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13351 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13401 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13451 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13501 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13551 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13601 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13651 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13701 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13751 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13801 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13851 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13901 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
13951 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
14001 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
14051 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
14101 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG  
14151 GAGGAGAGAG GAGCTGAGAG GAGCTGAGAG GCTGCTGAG GTGAGAGAG

FIGURE 3, page 4 of 22



[illegible]

FIGURE 3, page 6 of 22















GCTACAGGAGCGGTG \* AC TATGCCACAGCTTAATTTTGTATTTTATGTAAGAGAGAGGGTTT  
 TAAATATGTTGGTAA \* AGGTTTCTTAATTAATTAAGTATGAGTGGGTCATCTGCTGTCTAGGGGT  
 GCTAAATGTTATGAA \* TAAAGGCAATGGGATCTATGAGGAGGATGATGATATTAATATAGCA  
 AGGTTCATCTGTGT \* CAGAGGCTGAAGTAAAGAGCAATTAATAGTCTAAATGAGGCT  
 CAATATATCGAGT \* AGCGAGTACAGCTGAGTACAGTACAGTCAATATTTT

[illegible][illegible]

22375 ACAAACCCCTTCCGAGCCTATACATATGGCCACCTTGGAGCAAGAGAGGAGAGATATTTTC  
CCTGGGAGAGTCCTTTCAGCAAGATCTGGCCAGAGCTGGCCCTGCAGATCTCTTAAAGAAAGAG  
CAGCCCATGGCTTCCAGTGGGAGGAAGCATCTCACACTCTTTGGCAGAGAAAGATAGTTCGC  
AGGCCGCAGCTGTTGTGGAAGACCACTCGTTCTGTGGTTGGGGTCTGCGAAGAGAGGCTTCG





Isoform 1:

FEATURES:

Exon: 2137-2139  
 Intron: 2139-5570  
 Exon: 5571-5583  
 Intron: 5584-10102  
 Exon: 10103-10274  
 Intron: 10275-10399  
 Exon: 10400-10586  
 Intron: 10587-106129  
 Exon: 12129-12294  
 Intron: 12295-25922  
 Exon: 25923-26040

Allelic Variants (SNPs):

DNA				Protein		
Position	Major	Minor	Domain	Position	Major	Minor
1-2	C	T	Beyond ORF(5')			
1-42	C	T	Beyond ORF(5')			
1-1	G	A	Beyond ORF(5')			
2-10	A	G	Intron			
3-21	G	A	Intron			
6732	G	A	Intron			
7013	G	A	Intron			
7031	T	A	Intron			
8190	T	C	Intron			
8191	C	T	Intron			
8190	G	C	Intron			
8403	C	T	Intron			
8883	G	T	Intron			
10134	G	A T	Intron			
10161	A	G	Intron			
11187	T	C	Intron			
11484	T	C	Intron			
11541	C	G	Intron			
11796	A	G	Intron			
11799	C	T	Intron			
11804	C	A	Intron			
11946	C	-	Intron			
12476	C	T	Intron			
12143	A	- T	Intron			
12445	T	C	Intron			
12974	A	-	Intron			
14165	G	T	Intron			
15140	T	C	Intron			
15246	G	A	Intron			
15193	C	T	Intron			
15119	T	G	Intron			
15614	C	T	Intron			
15735	C	T	Intron			
16166	A	-	Beyond ORF(3')			

Context:

RNA

Position

1022 TTGGAGATATTTTAAAGGTCATACTGTCTTCACAAATTGAGCTGAAAGGGAAGTGTAGGA  
 TGATCTTGCCCTAAGCCTCTCATCTCACACAGGAAGAAGTATTTAAAGTCGAGAGGTTAA  
 GTGACCTGGCCAAAGTCACACAGCCACCACTAGTTAACTCGTATACATTGATTCTCCTGT  
 GGGGCTGGGCAGATGAGGAATCTTTTGTCTCTTCCCTGTTTGCAGAGATTTTGTGAG  
 GTTACTTTCCGAGTTCTGGCAAGTACCCCTGCTTCTGGTAGCTTGTGTCTCGATTCAT  
 [C,T]  
 TCATTCTTTTATTTTATTTTATTTTGTGAGACAGGCTCTCACTTTGTACCCAAAGTGGG

FIGURE 3, page 16 of 22



100

1951

2940

3831

6732

7558

FIGURE 3, page 17 of 22





23445 A G G A A A C T G G T T T T A A A A A G G A A A A G T G A C T C A C C A A G G T C A C A G A T A G G C A G T A T G  
C T G T G G G A A C C C G G C T A G G G G A C A C A G A C T G G C C T G G G G C A G C C T G C G A G T C C T C A  
C T A A A A T A C T G A A A A T C A G G G G C T T C G A T G A T G T T A T A A T C G F A T S C A G A G C C C C A A

GGGGGCTTAA GGGGCTCTGGGCTGAGCTTTTGAAGAGGGGGGTTCAAGCCGAGGAGGACAC  
CCTATGCTGTACCCGACCAAGTTAGGAAGAGGGTTCTTTTCAAGTGGGGGCTCTTAAG  
AAGACGCCAGCAGGCTTCTGTTCCAGATGCGTTGGGAAGAGATGGTGACATATCTAAATCT  
TTGAGACGGCATATTAAACACCTAGAGCAATCATCTTCAAGACCTTTAGGGCTGAGAA  
TAGGGGAAGAGCAATTTAAAGAGCTTTTATATTGGCGGAGGTGGCAATGGTTACCTT

[illegible][illegible][illegible][illegible][illegible]

FIGURE 3, page 21 of 22

26266  
 3TGAATGGGGAAGGGGGGCGGA AAGAGGAGTTCTGGCA GTTCTCGGAGACCCCAACACC  
 CCACCTGTTGGCAAGGTGAGTCATGGGCTGACACCTGGGA GGTGTCCCTTACCCCAAGCTTA  
 CTCAGGCCAAGAGGCTTCATCAAGTCACCCCAAGCTTTCC CTAGCACCCTCCCTGGGCCACAC  
 CTTTCACAAAATCACTGATGCTCAAGTTGGGATATAATA CATTGAACCTGAAGCCTTAGCAT  
 TTTTATGCAAGTTACTGTGGAANTTCATAGGAAACCAAGA TAGATTACAAAAAATAAAAAA  
 [A, -]  
 CTAGAAGAAAATTAACATCACCTAGGATATACTACCTA GGAATAACGTCCTTTTATTITGA  
 GATGGAGTTTCGCTCTCTGTTGTCAGGCTGGAGTGCAG GGTATGATCTTCGGCTCGCTGTC  
 AACCTCGGCTCTCTGGGTTCAATGATTTCTTCACCTC GCTCTCTCTAGAGGCCCAAGTGG  
 TCTGCTTCCTCTCTGCTCCCAAAGTTCTTGGATTACAG GATAGGCCACCGCACCCAGGC  
 AAAATTACTTTAACTTTTCTCTAGATATTTTTAAAAATATGGCAGTAGTTTCTTCATAA